Araújo, Isabel; Carneiro, Anabela

Working Paper

Educational Mismatches of Newly Hired Workers: Short and Medium-run Effects on Wages

GLO Discussion Paper, No. 668

Provided in Cooperation with:
Global Labor Organization (GLO)

Suggested Citation: Araújo, Isabel; Carneiro, Anabela (2020) : Educational Mismatches of Newly Hired Workers: Short and Medium-run Effects on Wages, GLO Discussion Paper, No. 668, Global Labor Organization (GLO), Essen

This Version is available at:
http://hdl.handle.net/10419/224274

Educational Mismatches of Newly Hired Workers: Short and Medium-run Effects on Wages*

Isabel Araújo

Universidade Lusíada, COMEGI

e-mail: isabelaraujo@por.ulusiada.pt

Anabela Carneiro

Universidade do Porto, CEF.UP

e-mail: anacar@fep.up.pt

September, 2020

Abstract

Exploring a rich matched employer-employee data set over the 1998-2012 period and a novel measure of educational mismatch, this study analyses the short and medium-term effects of over- and undereducation on the wages of newly hired workers. The data show that more than 50 percent of the employed in the private sector in Portugal experienced a job mismatch at the moment of being hired. According to the statistical measure based on the flows of newly hired workers, in the

*We acknowledge Gabinete de Estratégia e Planeamento - Ministério do Trabalho, Solidariedade e Segurança Social (GEP-MTSSS) for allowing the use of *Quadros de Pessoal* dataset.This research has been financed by Portuguese public funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., in the framework of the projects with references PTDC/CED-EDG/29726/2017 and UIDB/04105/2020.

period under scrutiny overeducation is decreasing and undereducation is increasing, indicating that labour market demand is keeping pace with the rise in educational attainment of the Portuguese population.

The results reveal that the wage differential between adequately matched workers and mismatched workers decreases considerably once worker and firm unobserved heterogeneity is taken into account. In fact, worker permanent heterogeneity explains two-thirds of the overducated wage penalty and three-fourths of the undereducated wage premium, indicating that the undereducated seem to correspond to a higher-ability group of employees, while the overeducated seem to correspond to a lower-ability group of workers. Heterogeneity in firm paying policies also play an important role in explaining the wage gap of newly hired mismatched workers.

Finally, the results also indicate that the wages of individuals in the beginning of their labour market career are the most affected by a job mismatch.

KEYWORDS: educational mismatches, overeducation, undereducation, wages, two-way fixed effects

JEL CODES: I26; J24; J31

1 Introduction

Skill mismatches in the labour market are a matter of great concern for academics, practitioners, and policymakers. Skill mismatches arise from several imbalances between the skills offered and the skills demanded in the labour market due to search and matching frictions such as information asymmetries, mobility costs, and education and training systems that do not fit the labor market needs (Quintini, 2011; ILO, 2014). For firms, these imbalances lead to inefficiencies in the utilization of labor, with detrimental effects on productivity (Tsang and Levin, 1985; Kampelmann, Mahy, Rycx and Vermeylen,

2020) and turnover rates (Sicherman, 1991; Hersch, 1991). For workers, they can affect job satisfaction (Tsang, Rumberger, and Levin, 1991; Maynard, Joseph, and Maynard, 2006; Erdogan, Bauer, Peiró, and Truxillo, 2011), training investments (Groot, 1993; Van Smoorenburg and Van der Velden, 2000) and wages (Hartog, 2000; Groot and Van den Brink, 2000). Positive outcomes of overqualification associated with training and job performance are also documented in a few studies (e.g., Fine, 2007; Fine and Nevo, 2008).

In a recent survey, McGuiness, Pouliakas, and Redmond (2018) focusing on published papers that analyse skill mismatches over the 2006-2016 period, provided estimates of the incidence of over- and undereducation based on the realized matches approach. According to the estimates of 103 (23) papers, the authors report an overall average incidence of overeducation (undereducation) that reaches 25.9 (26.2) percent, highlighting the relevance of the phenomenon even though uncovering some heterogeneity across countries.

Previous studies investigated the earnings consequences of over- and undereducation augmenting the standard mincerian wage equation with measures of over- and underschooling (Duncan and Hoffman, 1981; Rumberger, 1987; Verdugo and Verdugo, 1989; Sicherman, 1991; Cohn and Kahn, 1995 for the U.S; Hartog and Oosterbeek, 1988 for the Netherlands; Alba-Ramirez, 1993 and Murillo, Rahona-López, and Salinas-Jiménez, 2012 for Spain; Kiker, Santos, and Oliveira, 1997, and Oliveira, Santos, and Kiker, 2000 for Portugal; Cohn and Ng, 2000 for Hong Kong; Dolton and Vignoles, 2000 for the U.K.). This first wave of empirical studies, drawn on cross-section data, established two basic stylized facts. First, based on the Duncan and Hoffman (1981) model, the empirical results show that, when compared with their job co-workers who are adequately educated, overeducated workers receive a wage bonus for the extra years of surplus schooling and undereducated workers a wage penalty for the deficit years of schooling, even though

smaller than the returns to required education. Second, based on the Verdugo and Verdugo (1989) model, the empirical evidence showed that overeducated workers earn less, and undereducated workers earn more than their counterparts with the same years of schooling, but who hold jobs for which they are adequately educated.

Overall, these findings seem to support the hypothesis that wages are not uniquely determined on the basis of the individual's educational level. The characteristics of the job seem also to play an important role in wage determination favouring an assignment theory interpretation of the labour market (Sattinger, 1993). According to the assignment theory, productivity depends on both - worker and job characteristics - and is maximised when workers are allocated to the jobs according to their skill ranks. Most skilled workers should be assigned to the most complex and demanding jobs, whereas the least skilled to the simplest jobs. In this setup, workers with the same attained level of education may perform differently depending on the job they hold, allowing individuals to make their choices according to their relative comparative advantage in the possible set of job offers.

In the last two decades, a second wave of studies, drawn on longitudinal data, further explored the wage effects of educational mismatches. Overall, this literature showed that the wage gap between well-matched and mismatched workers reduces substantially, once unobserved individual heterogeneity is taken into account (see Bauer, 2002 for Germany; Frenette, 2004 for Canada; Tsai, 2010 for the U.S.; Mavromaras, McGuinness, O'Leary, Sloane, and Wei, 2013 for Australia), suggesting the existence of a trade-off between over- and undereducation and other components of human capital (Sicherman, 1991).

The purpose of this paper is to extend this recent literature by exploring a novel measure of educational mismatch based on the flows of newly hired workers. Acknowledging that statistical measures of educational mismatches based on the mean or mode of the

stock of employees per occupation suffer from insurmountable drawbacks covering up a great heterogeneity in the pool of employees who were probably hired in very different time periods (see Kiker *et al.*, 1997 for a detailed discussion on this issue), this study proposes a novel measure of required education based on the flows of newly hired workers within a 3-digit occupation. Then, to evaluate the effect of educational mismatches on wages we compare the wage changes of workers who were hired mismatched with the wage changes of similar workers who were hired well-matched, controlling, in addition to worker unobserved heterogeneity, to the role of firm observed and unobserved permanent heterogeneity to account explicitly for demand side variables.

Recruitment and selection entail a two-sided matching process between the firm and the worker that are simultaneously determined by observed and unobserved characteristics of both parties. The reasons why an employer may decide to hire a worker that is apparently underqualified for a given job or why a worker may accept a job for which his/her attained level of education exceeds the level required for the job may be driven by unobserved characteristics of the worker and the employer. Failure to account for endogeneity in these choices may bias the estimates of educational mismatches on wages.

Thus, in this study we acknowledge at least two endogeneity problems that should be addressed in our empirical methodology. The first stems from the fact that individuals are heterogenous and differ in certain unobserved characteristics. If these unobserved characteristics are correlated with the mismatch status, neglecting them may bias the estimated effects of educational mismatches on earnings (Chevalier, 2003; McGuinness, 2006). Actually, some studies found evidence that ability and overschooling are negatively correlated (e.g., Greene and McIntosh, 2007; Chevalier and Lindley, 2009).

The second potential source of endogeneity may arise from selection effects since work-

ers sorting across firms is non-random (Baptista, Lima, and Preto, 2013; Rocha, van Praag, Folta, and Carneiro, 2019). In this framework, workers self-selection imply that individuals make choices regarding whether they join a given firm/job based on their own attributes and firm/job characteristics (most often not observed by the researcher) that may be correlated with both educational mismatches and expected wages. Suppose that workers are more likely to accept a job for which his/her level of education exceeds the required for the job in large/high-paying firms that offer more career prospects and a higher wage growth. In this case, ignoring firm heterogeneity may bias the estimates of the returns to overeducation as high-paying firms may be more likely to attract workers that are willing to accept a job for which they are overeducated. Employers selection is also a concern in this study. Recent studies on new ventures performance, have shown that hiring decisions are not independent from founders observed and unobserved characteristics (Rocha *et al.*, 2019). Suppose that business-owners or managers less qualified, risk averse or with low levels of confidence may be more likely to recruit candidates who have excess education. Neglecting these issues, potentially correlated with job mismatches and expected outcomes, may mislead the causal interpretation of the effects of over- and undereducation on wages.

We are confident that Portugal constitutes an interesting case to develop this exercise. First, Portugal experienced, in the past two decades, a huge increase in the educational levels of its population, favoured by the massive expansion of the higher education system (Figueiredo, Teixeira, and Rubery, 2013). Nevertheless, and despite the significant improvement in the educational levels of the labour force, Portugal still ranks below the OECD and UE averages in terms of schooling attainment (OECD, 2017). Second, the Portuguese labour market is characterized by very strict employment regulation legisla-

tion, where individual and collective dismissals of workers in permanent contracts are the most restrictive across the OECD (OECD, 2012; Martins, 2009; Centeno and Novo, 2012). The termination of a permanent contract in Portugal involves a lengthy and complex administrative procedure that imposes several costs to firms and creates barriers against workforce adjustments (Martins, 2009) that potentially difficult match quality improvements. Third, Portugal seems to exhibit, since the mid 1990s, a phenomenon of job polarization in the labor market mainly driven by technological changes (Fonseca, Lima, Pereira, 2018) that enhances the recruitment of better-educated cohort of workers. Using *Quadros de Pessoal* data for 1986-2007, Fonseca *et al.* (2018) show a sharp increase of employment in abstract tasks relative to manual tasks, along with a decline for routine manual tasks.

The contribution of this study to the current literature is threefold. First, to document the incidence of over- and undereducation in Portugal in the 1998-2012 period and its recent time trends, exploiting a novel measure of required schooling based on realized matches for the flows of new hires. Second, to complement prior research that aimed to estimate the economic returns to over- and undereducation by taking simultaneously into account the role of worker and firm unobserved heterogeneity. Third, to analyse the short- and medium-run effects of educational mismatches on wages to shed further light on the temporary (or not) nature of the phenomenon.

Our contribution relies on two main ingredients: (i) a rich administrative longitudinal matched employer-employee data set, and (ii) an identification strategy that allows to simultaneously account for worker and firm unobserved heterogeneity to properly address endogeneity in job mobility and educational mismatch status.

The rest of the paper is organized as follows. Section 2 describes the data and reports

figures of the incidence of over- and undereducation in Portugal in the 1998-2012 period. Section 3 describes the econometric model and discusses identification issues. Empirical results and robustness checks are presented and discussed in Section 4. Section 5 concludes.

2 Data and Methodological Issues

2.1 Quadros de Pessoal (QP)

Our data come from *Quadros de Pessoal* (*QP*), a matched employer-employee dataset collected by the Portuguese Ministry of Labor, Solidarity, and Social Security. *QP* is an annual mandatory survey that all Portuguese firms in the private sector with at least one wage earner are legally obliged to fill in.[1] Data are available from 1985 until 2018 and include information at the firm, establishment, and worker level.[2] At the firm level, *QP* contains information on industry, location, employment, sales, ownership, legal setting, among others. Worker data include information on gender, age, education, occupation, qualification, tenure, wages, hours worked, among others.

Firms, establishments, and workers entering the database are assigned a unique identifying number that makes it possible to track them across all annual waves of data. Furthermore, the worker files include the firm and establishment number to which each individual is affiliated in a given year, allowing to match workers with their employers both at the firm and the establishment level.

Our data covers the population of employed workers in the private sector, ignoring the unemployed, including those in searching for the first job. Some authors argue that ignor-

[1] Public administration, self-employment and nonmarket services are not covered by *QP*.
[2] Worker data are not available for the year 2001.

ing the non-employed might generate a sample selection bias as those who are employed may constitute a non-random sample of the population with unobserved characteristics that may be correlated with the educational mismatch (Cutillo and Di Pietro, 2006; Caroleo and Pastore, 2016). However, the longitudinal nature of the dataset, the long time span covered and its high degree of representativeness and reliability, makes *QP* an appropriate database for a comprehensive study on the wage effects of educational mismatches of newly hired workers, making selection problems driven by the lack of information on the unemployed less that an issue in our case.

Finally, employer reported information is, in general, subject to less measurement error than worker reported information.

2.2 Measuring Educational Mismatches

The measure used to define educational mismatches among the employed is crucial to our analysis and previous literature showed that the patterns of skill mismatches strongly depend on the criteria adopted to measure it (e.g., ILO, 2014; Groot *et al.*, 2000). In this paper we will focus on vertical educational mismatches in the Portuguese labor market. A vertical mismatch takes place when the level of education is higher or lower than the one required for the job. Following previous studies, we will rely on a statistical measure based on realized matches (e.g., Verdugo and Verdugo, 1989; Oliveira *et al.*, 2000; Hartog and Groeneveld, 2004; Korpi and Tåhlin, 2009; Bauer, 2002). Previous studies evaluated required education based on the mean or mode education of the stock of employees in a given occupation. As pointed out by McGuiness *et al.* (2018) "Drawbacks of the realized matches method are that it does not contain any information on the actual skill requirements of the job, it reflects average credentials of all workers within a given

occupation and, therefore, is more closely related to education levels required 'to get' as opposed 'to do' a given job in contemporary terms." To overcome this drawback, our measure, a novelty in this study, takes more properly into account that the level of required education may change over the years as employers hiring standards adjust to technological changes, organizational structure changes, increases in the relative supply of higher educated workers, etc. Acknowledging that what really matters most are entry requirements at the moment of being hired, our measure of over- and undereducation will be based on realized matches of newly hired workers. In this case, required education is defined as the mode level of education for the flow of newly hired workers (tenure less than 12 months) in a 3-digit occupation in a given year.

To apply the statistical measure, first we need to convert the completed levels of schooling available in *QP* in a quantitative variable measured in years of schooling. The Portuguese educational system is structured in three levels: primary (first cycle, second cycle, third cycle), secondary (secondary and post-secondary), and tertiary education (bachelor, graduate, master, doctoral degree). We assigned years of schooling as follows: less than first cycle - 0 years; first cycle - 4 years; second cycle - 6 years; third cycle - 9 years; secondary education and post-secondary education - 12 years; bachelor - 15 years; graduate, master degree, doctoral degree - 16 years.[3]

As the Portuguese Classification of Occupations (CPP) changed in 2010, the occupation codes valid before 2010 were recoded according to the new classification of occupations. For a few occupations it was not possible to accurately match the codes and, thus, these occupations were excluded from our analysis. Finally, and to accurately identify required education, occupations that represented less than 0.01% of the total number of

[3]Given that the distinction between graduate, master, and doctoral degree is reported in *QP* only after 2005 and to use a homogeneous criterion over the entire period of analysis, we decided to assign 16 years of schooling for graduates, masters, and doctorates.

newly hired workers in a given year were also dropped. We end up with 92 occupations at a 3-digit level.

Based on this definition of required education and in order to classify the individuals as over- or undereducated, required education (RE) for a given 3-digit occupation is compared to the current level of schooling attained (AE) by the worker in that same occupation.[4] Thus, in a given time period each individual is classified in one of three groups:

(i) Overeducated (OE): if $AE > RE$;

(ii) Undereducated (UE): if $AE < RE$;

(iii) Matched (M) - base category: if $AE = RE$.

2.3 Sample Construction

As explained before, this study aims to evaluate the short and medium-run effects on wages of educational mismatches when the individuals were first hired for the job. Thus, we restricted our sample to workers who were hired at least once between 1995 and 2012 in mainland Portugal aged between 17 and 65 years old.[5] Workers are then tracked for three consecutive periods after the entry year. Thus, the 1995, 1996, and 1997 waves are only used to characterize the educational status (overeducated, undereducated, and matched) of the worker in subsequent years. For example, for a worker that appears in the QP files in year 1998 with three (two) years of tenure the information of year 1995 (1996) is crucial

[4]For a detailed discussion of the advantages and limitations of the methods used to measure educational mismatch see, for example, Kiker *et al.* (1997), Hartog (2000) and McGuiness *et al.* (2018).

[5]Information on occupations from 2013-2018 is provided at the 4-digits level instead of the 6-digits level, which precludes us to properly recode the occupations according to the 2010 change in the Portuguese Classification of Occupations - a crucial variable in our study.

to identify his/her educational status in the entry year in the current firm. After being assigned to one category (OE, UE, or M), the worker remains in that same group until he/she moves to a different employer.[6] Should the worker move to a different firm, his/her individual clock resets to zero and his/her assignment to the initial group is cancelled and reassigned to the new group of educational mismatch.[7] It should be noticed, that some individuals may change his/her educational status within the same firm. Actually, a switch in the mismatch status within the same firm takes place if required education (RE) varies as individuals switch occupation within firms or employers' hiring standards change over time within the same firm-occupation. Furthermore, the educational status of the individual in a given firm may change if the attained level (AE) of education varies overtime as individuals invest on formal education while participating in the labor market. To adequately date the worker's educational status at firm entry, we only used information on required and attained education at the moment of being hired (tenure less than 12 months), ignoring changes in educational status that ocurred during the course of the employment relationship.

The data include 7,204,555 (years×individuals) observations of workers employed over the 1998-2012 period and with a maximum of three years of tenure, corresponding to 3,393,381 newly job matches, 1,880,550 individuals and around 249,437 thousand firms.[8] Table 1 reports the number of observations by type of educational mismatch at job entry.

INSERT TABLE 1 HERE

[6] An employer change occurs when the firm's identifying number in year t and $t-1$ is not the same and tenure of the worker in year t is less than 12 months.

[7] As data on workers are not available for the year 2001, worker's educational status cannot be defined for those individuals that were hired in 2001. Thus, these observations were dropped. Observations with missing values on the variables of interest were also excluded.

[8] For estimation purposes, singletons (individuals with only one observation) were excluded as well as observations for which it was not possible to identify the firm fixed effect.

2.4 Incidence of Over- and Undereducation

Table 2 reports the average percentage of overeducated and undereducated workers in Portugal for the 1998-2012 period. Our results show that around 50 percent of the Portuguese workers in the private sector experience an educational job-mismatch at the moment of being hired. The average percentage of overeducated is 30 percent, while the average percentage of undereducated is 23.4 percent. These percentages are comparable with the average values shown in the recent survey conducted by McGuiness *et al.* (2018), who reported an average of 25.9 (26.2) percent of overeducated (undereducated) drawn from 103 (23) data-based estimates of a list of published papers. Regarding the incidence of the phenomenon by gender, unconditionally, women are more likely to be adequately educated than men. In the 1998-2012 period, the average percentage of males that are overducated (undereducated) is 31.6 (23.9) and the average percentage of females 27.5 (22.6).

INSERT TABLE 2 HERE

Table 3 reports the incidence of over- and undereducation by occupations at the 1-digit level for newly hired workers in Portugal (average of the 1998-2012 period). The data reveal that only three occupations have more than half of the total employed in that occupation well-matched, namely "Professionals", "Clerical support workers", and "Skilled agricultural, forestry and fishery workers". "Managers" is the occupation that presents a larger proportion of undereducated workers (about 36 percent), while "Elementary occupations", "Skilled agricultural, forestry and fishery workers", and "Craft and related trade workers" are the occupations that present a larger fraction of overeducated individuals.

INSERT TABLE 3 HERE

Figure 1 shows the evolution of the incidence of over- and undereducation over the analysed period, while Figure 2 presents the evolution of the average years of schooling for the newly hired workers. Overall, the data reveal a decreasing trend in overeducation and an increasing trend in undereducation.[9] In fact, the percentage of undereducated workers increased from 18.7 percent in 1998 to 26.1 percent in 2012, while the percentage of the overeducated decreased from 32.7 percent in 1998 to 25.6 in 2012 (Figure 1).

These patterns are consistent with a substantial increase in the levels of educational attainment experienced by the Portuguese population in the last two decades and an upgrade of the recruitment educational standards of labor demand. The average number of years of schooling for the newly hired increased from 7.5 years in 1998 to 9.9 years in 2012 (Figure 2).[10] The data suggest that the labour market demand kept in pace with the rising in the educational attainment of the Portuguese population in the past two decades.

INSERT FIGURES 1 AND 2 HERE

3 Econometric Model

The purpose of this Section is to evaluate the impact of educational mismatches on wages comparing the wage changes of workers who were hired mismatched ("treatment group") with the wage changes of similar workers who were hired matched ("control group"). Table 4 presents the average hourly wages by educational mismatch status in the baseline year and the three subsequent years. Hourly wages correspond to total regular payroll (base

[9] The trend reversal in 2012 may be explained by the 2011 Portuguese debt sovereign crisis that led to a massive job destruction (Carneiro, Portugal, and Varejão, 2014) as overeducation is more likely to occur during recessions (Clark, Joubert, and Maurel, 2017).

[10] Of the total number of 92 occupations at the 3-digit level, 50 increased its required level of education between 1998 and 2012.

wage and regular payments) over normal hours worked in the reference month.[11] Hourly wages are, on average, higher for the matched workers and lower for the undereducated employees. Real hourly wages for the overeducated lie between the averages of the former and the latter group of workers, respectively.

INSERT TABLE 4 HERE

As in previous research, we acknowledge that observed characteristics of the individual such as education and experience imperfectly reflect their true productivity (Abowd, Kramarz, and Margolis, 1999; Iranzo, Shivardi, Tosetti, 2008). Unobserved attributes such as ability, motivation, resilience, play an important role in wage determination that, if ignored, lead to biased OLS estimates. Furthermore, we claim that mobility across firms in searching for a well-matched job raise endogenous problems, whereas sorting based on unobserved characteristics of both sides of the labor market - workers and firms - is likely to occur (e.g., Baptista *et al.*, 2013; Dahl and Klepper, 2015; Rocha *et al.*, 2019).

In the spirit of Verdugo and Verdugo (1989), the econometric model that we use is a standard mincerian wage equation augmented to include controls for educational mismatch at firm entry accounting for worker observed and unobserved (permanent) heterogeneity and firm observed and unobserved (permanent) heterogeneity. The model writes as:

$$\ln w_{ijt} = \alpha_i + \theta_j + \gamma_t + \boldsymbol{\beta X}_{ijt} + \sum_{k=0}^{3} \delta_k OE_{ijt}^k + \sum_{k=0}^{3} \lambda_k UE_{ijt}^k + \varepsilon_{ijt} \tag{1}$$

where w_{ijt} represents the real hourly wages for each individual i, employed at firm j in year t. α_i is a worker fixed effect, θ_j corresponds to the firm fixed effect, and γ_t a time fixed effect. Following Jacobson, Lalonde, and Sullivan (1993) and Tavares, Carneiro, and

[11] Wages were converted into 2010 constant prices using the Consumer Price Index (CPI).

Varejão (2018), OE_{it}^{k} and UE_{it}^{k} are dummy variables that take the value one if at time t, worker i is k years ($k \in [0,3]$) after being hired in firm j into a 3-digit occupation where he/she was overeducated or undereducated at the moment of being hired, respectively.[12] Vector $\boldsymbol{X}_{ijt}$ includes the attained level of education (in years), a quadratic term in age and tenure, and a set of dummies for qualification level (Table 5 presents the descriptive statistics of all the variables included in the model. ε_{ijt} is a random error term assumed to be uncorrelated with the regressors.

Vector $\boldsymbol{X}_{ijt}$ includes the attained level of education (in years), a quadratic term in age and tenure, and a set of dummies for qualification level (Table 5 presents the descriptive statistics of all the variables included in the model). ε_{ijt} is a random error term assumed to be uncorrelated with the regressors.

The parameters of interest are δ_k and λ_k that capture the wage differential of over- and undereducated workers, respectively, to the omitted category (similar matched workers), not only in the year of hire but also in the subsequent three years at the current firm, allowing us to capture the short and medium-term effects of over- and undereducation. According to the human capital theory (Becker, 1964), if worker productivity is not affected by the job requirements and years of schooling are a good proxy for productivity, the δ_k and λ_k coefficients should be zero as firms adjust wages to worker's marginal productivity. If productivity depends on the interaction between job characteristics and worker characteristics as argued in the job assignment interpretation (Sattinger, 1993), workers with the same educational level may perform differently depending on the job they are doing. In this context, δ_k and λ_k coefficients should be different from zero.

[12] For example, consider a worker that was hired in 1998 overeducated and then changed employer in 2000 to a job where he/she remains overeducated. Thus, in 1998 (the baseline year) the dummy variable OE_0 is equal to one, zero otherwise; in 1999 the dummy variable OE_1 is equal to one, zero otherwise; in 2000 the worker changes employer and his/her individual clock resets to zero and, thus, the dummy variable OE_0 equals to one once again.

In a fixed effects approach, identification of the parameters of interest comes from individuals that changed their educational status in the period under scrutiny. Our data show that about one-third of the individuals experienced a change in his/her job match status in the 1998-2012 period corresponding to 34.8 percent of the total number of observations.[13]

INSERT TABLE 5 HERE

4 Empirical Results

4.1 The magnitude of the over- and undereducated wage gap

Table 6 reports the estimated returns to educational mismatches based on the full specification defined in equation (1). For comparison purposes, we also provide the estimates of two additional specifications. The first, reported in column (1), was estimated by pooled OLS (labeled "base model"). The second specification, reported in column (2), is augmented to account for worker fixed-effects (FE). Finally, the specification in column (3) includes both worker and firm fixed-effects (labeled "full model").[14]

The OLS estimates of the returns to over- and undereducation reported in column (1) are consistent with earlier literature.[15] Overeducated workers earn less and undereducated workers earn more than their counterparts with the same attained level of education who

[13] Of this total of 34.8 percent of observations, 4.8 percent correspond to individuals that were at least once overeducated and undereducated in the analysed period, 14.7 percent correspond to individuals that were at least once overeducated and well-matched, 11.5 percent to individuals that were at least once undereducated and well-matched and, finally, 3.8 percent of the observations correspond to individuals that experienced the three status (overeducated, undereducated, and well-matched) in the period under scrutiny.

[14] The high-dimensional fixed effects models were estimated using the "reghdfe" stata command developed by Correia (2016). For sake of parsimony, only the estimated coefficients for OE^k and UE^k are reported in the Tables. Full results are available upon request.

[15] The signs of the coefficients of the control variables are consistent with the usual findings: wages increase at a decreasing rate with experience (proxied by age) and tenure, and are lower for female workers, the less educated and the less skilled workers.

work in jobs that require their current level of education at the moment of being hired. For the overeducated, the wage penalty ranges from around 7.3 percent in the year of hiring to 4.7 percent three years after, suggesting that the gap tends to vanish as the job match evolves.[16] For the undereducated, the wage premium ranges from 5.8 percent in the year of hiring to 4.5 percent three years after.

The worker FE estimates reported in column (2) show a significant reduction in the wage gap between mismatched and matched workers once unobserved individual heterogeneity is taken into account, corroborating previous studies using panel data (e.g., Bauer, 2002; Frenette, 2004; Korpi and Tahlin, 2009; Tsai, 2010; Mavromaras *et al.*, 2013). Actually, now entering into a job for which the individual is overeducated decreases wages by 1.7 percent in year of hiring and by 0.9 percent three years after, while entering undereducated increases wages by solely 0.8 percent in the year of hiring and by 0.5 percent three years after.

Finally, once firm permanent observed and unobserved heterogeneity is accounted for, the wage gap between the overeducated (undereducated) and the matched workers is further reduced (slightly increased). Furthermore, the estimates in column (3) exhibit a positive trend in wages for the overeducated as the wage penalty changes from -1.6 percent in year zero to -0.8 percent three years ahead. For the undereducated the tiny wage premium remains almost stable over the years.

INSERT TABLE 6 HERE

At a first glance, it seems that accounting for firm permanent heterogeneity does not affect substantially the wage gap between mismatched and matched workers. However, a simple comparison across specifications of the estimates obtained for the coefficients

[16] The exact values of the gap are computed as $(exp(\widehat{\delta} \text{ or } \widehat{\lambda}) - 1) * 100$.

of interest may be misleading, due to sequence sensitivity when added covariates are correlated (Gelbach, 2016). In our case, the relationship between the mismatch dummies and the worker and the firm fixed effect may be sensitive to the order in which these regressors are added to the model. This issue is particularly relevant if worker and firm fixed effects are correlated as workers sorting across firms is non-random (Baptista, Lima, and Preto, 2013; Rocha, van Praag, Folta, and Carneiro, 2019).

To evaluate the independent contribution of each fixed effect to the change in the wage gap of mismatched workers, we use the Gelbach decomposition (Gelbach, 2016) that appeals to the omitted variable bias formula. Beginning with a baseline specification ("base model") to which covariates are added ("full model"), Gelbach's procedure allows us to compute the individual contribution of each new covariate to the change in the regression coefficients of interest. In our case, it allows us to unambiguously disentangle the contribution of each excluded variable (each fixed effect) to the variation of the coefficient estimates of the educational mismatch dummies. Table 7 displays the results of the Gelbach decomposition.

Regarding the overeducated, conditional on education, age, tenure, gender dummy, qualification and time dummies, worker permanent heterogeneity accounts for nearly two-thirds of the difference between the estimates in specification 1 and in specification 3, while firm fixed effects for the remaining one-third. For example, considering the baseline year, of the total change of -5.8 log points in the educational mismatch dummy OE_0, -3.3 log points are due to the worker fixed effect, while -2.4 log points are due to the firm fixed effect. These results indicate that a large fraction of the wage penalty for the overeducated is explained by worker permanent (observed and unobserved) characteristics that are associated with lower wages. Additionally, overeducated workers earn lower wages in part

because the firms from which they were hired exhibit a less generous wage policy.

Regarding the undereducated, worker permanent heterogeneity accounts for around three-fourths of the difference in the wage premium between specifications 1 and 3, while firm fixed effects for the remaining one-fourth. For example, looking at the baseline year, of the total change of 5.0 log points in the educational mismatch dummy UE_0, 3.7 log points are due to the worker fixed effect, while 1.3 log points are due to the firm fixed effect. These results indicate that a large fraction of the wage premium earned by the undereducated is explained by permanent (observed and unobserved) characteristics of the worker that are associated with substantially higher wages. Additionally, undereducated workers earn higher wages in part because, conditional on all covariates, the firms from which they were hired exhibit a more generous wage policy.

Overall, these results indicate that the effects of vertical educational mismatches on wages may be largely due to omitted variable bias and self-selection, thereby underlying the need to properly control for unobserved individual heterogeneity. Controlling for heterogeneity in firm paying policies, in addition to worker heterogeneity, is also relevant.

INSERT TABLE 7 HERE

4.2 The Empirical Distribution of Worker and Firm Permanent Heterogeneity

Figures 3 and 4 display the empirical distribution of permanent worker observed and unobserved heterogeneity for the samples of overeducated, undereducated, and adequately educated workers. Worker permanent heterogeneity is proxied by the estimates of the worker fixed effect obtained by the estimation of the full model defined in equation (1). The graphs are based on the 1,880,550 estimates of worker fixed effects filtered from firm

unobserved heterogeneity. A worker with a higher fixed effect is an individual with a higher earnings premium after controlling for his/her observed attributes, including the eventual educational mismatch, and firm observed and unobserved permanent heterogeneity. It approximates the individual's unobserved time-invariant attributes that are likely to affect productivity and wages.

Figure 3 provides the distribution of the worker fixed effects estimates for the overeducated against the well-matched, indicating clearly that the empirical distribution of the worker fixed effect for matched workers is more shifted to the right. Furthermore, the empirical distributions of worker permanent heterogeneity for the undereducated against the matched workers displayed in Figure 4, show that the former have unobserved permanent characteristics that are associated with higher wages when compared with similar matched counterparts. In other words, undereducated workers seem to correspond to a higher-ability group of employees, while overeducated workers seem to correspond to a lower-ability group of employees.[17]

Figures 5 and 6 depict the empirical distributions of firm permanent observed and unobserved heterogeneity. The graphs are based on the 249,437 estimates of firm fixed effects of the full model defined in equation (1). According to Figure 5, the distributions of the firm fixed effects for the overeducated and the matched workers exhibit a similar shape. Regarding the characteristics of the firms where the undereducated are employed, Figure 6 seems to suggest that the undereducated are employed in firms that, on average, pay low wages when compared with the matched workers, which seems to be inconsistent with the results obtained by the Gelbach decomposition.[18] However, when we partial out

[17] The Hotelling's T-squared test rejects the null hypothesis that the mean of the worker fixed effects are equal for the three groups considered (OE, UE, and, M).

[18] The Hotelling's T-squared test rejects the null hypothesis that the mean of the firm fixed effects are equal for the three groups considered (OE, UE, and, M).

all other variables and, in particular, the attained level of education, we conclude that the undereducated are, on average, employed in high-paying firms and the overeducated in low-paying firms when compared with similar individuals with the same level of schooling.

Overall, this evidence reinforces the hypothesis that overeducation may emerge as a mechanism to compensate any relative disadvantage in terms of skills, while undereducation may coexist if individuals with low levels of formal education compensate this disadvantage with other forms of human capital relevant to perform the job (e.g., ability).

INSERT FIGURES 3 TO 6 HERE

4.3 Robustness Checks

As a first robustness check, we estimate the models using an alternative measure of hourly wages. Table 8 presents the estimates of the three specifications reported in Table 6 but adopting as the dependent variable the natural log of the real base wage (excluding regular payments). Overall, the results are qualitatively similar to the ones reported in Table 6.

INSERT TABLE 8 HERE

As a second robustness check, and in order to evaluate how the wage gap of mismatched workers evolves as labor market experience increases and, on one hand, individuals invest on-the-job training and, on the other hand, reveal their true productivity skills to employers, we run the three specifications reported in Table 6 including interaction terms between the mismatched dummies and a dummy variable for labor market experience - labelled "exp". This dummy variable takes the value one for those individuals with more than 5 years of experience in the labor market ($exp = 1$ if $age - educ - 6 > 5$; 0 otherwise).

The estimates reported in Table 9 show that at labor market entry (exp $\leq$ 5) mismatched workers suffer a penalty whether they were hired overeducated or undereducated. This penalty tends to vanish as the job match evolves and workers' labor market experience increase. For the undereducated with more than five years of experience we observe a wage premium when compared with their similar matched counterparts. These evidences reinforce the idea that mismatched status may result from the lack of other human capital components, such as on-the-job training, and employers' asymmetric information regarding the true productivity of workers. Thus, as labor market experience increases and the job match evolves, these imbalances seem to become less relevant for those whose job matches survive.

INSERT TABLE 9 HERE

5 Concluding Remarks

Exploring a representative administrative dataset that covers the population of wage earners in the private sector in Portugal and a novel measure of educational mismatches, this paper provides a comprehensive analysis on the incidence of over- and undereducation of newly hired workers and its short and medium-term effects on individual wages. The empirical findings emerging from this exercise are fourfold.

First, the data indicate that half of the employees in the private sector in Portugal experience a vertical educational mismatch in the moment of being hired. Overeducation is more likely among male than female workers.

Second, a time trend analysis over the 1998-2012 period revealed that overeducation is decreasing and undereducation is increasing as a consequence of the upgrade in employers

hiring standards that keep in pace with the substantial rising in the educational attainment of the Portuguese population in the past two decades.

Third, accounting simultaneously for worker and firm permanent observed and unobserved heterogeneity reduces substantially the returns to over- and undereducation, suggesting that educational mismatches are largely driven by unobserved characteristics of the worker and the firm and failure to account for them bias the estimates of the mismatch educational effects. Based on the omitted variable bias formula proposed by Gelbach, we found that worker permanent heterogeneity explains two-thirds of the overeducated wage penalty and three-fourths of the undereducated wage premium, indicating that the undereducated seem to correspond to a higher-ability group of employees, while the overeducated seem to correspond to a lower-ability group of workers. Firm heterogeneity also plays a role in explaining the wage gap between mismatched and matched workers. Conditional on covariates, overeducated are employed in low-paying firms when compared with their matched counterparts, while the undereducated are employed in high-paying firms when compared with their similar matched workers.

Fourth, and finally, we found that workers in the beginning of their labor market career are the most affected by an educational mismatch whether they are over- or undereducated. As labor market experience increases and employers have access to more information on workers' skills and workers invest on-the-job training, the wage penalty suffered by the over- and undereducated tends to vanish (becoming even positive for the undereducated) for those that were able to keep their jobs, highlighting the importance of analysing the short- and medium-term effects of the phenomenon.

Overall, this paper contributes to the literature on the wage effects of vertical educational mismatches by proposing a novel measure of over- and undereducation based on

the flows of newly hired workers. We also emphasize the endogeneity issues driven by workers' and employers' selection in educational mismatch status using a methodological approach that addresses endogeneity driven by worker and firm unobserved heterogeneity. This issue is particularly important as in the last years the pool of individuals with higher education became more heterogenous in terms of their unobserved characteristics with the massification of the access to tertiary education (Bauer, 2002; Cutillo and Di Pietro, 2006). Finally, to evaluate to what extent over- and undereducation have long lasting effects we adopted a model's specification that allows us to isolate the immediate and medium-term effects of educational mismatches.

To complement this analysis, future research should focus on the costs of educational mismatches in terms of unemployment and future employment prospects as in our approach we were not able to evaluate to what extent previous unemployment spells impact on mismatch status and vice-versa.

References

[1] Abowd, J. M., Kramarz, F., Margolis, D. (1999). "High wage workers and high wage firms". Econometrica, 67(2): 251-334.

[2] Alba-Ramirez, A. (1993). "Mismatch in the Spanish labor market: Overeducation?". Journal of Human Resources, 27(2): 259-278.

[3] Bauer, T. K. (2002). "Educational mismatch and wages: a panel analysis". Economics of Education Review, 21: 221-229.

[4] Baptista, R., Lima, F., Preto, M. T. (2013). "Entrepreneurial skills and workers' wages in small firms". Small Business Economics, 40(2): 309-323.

[5] Becker, G. (1964). "Human Capital: A theoretical and empirical analysis with special reference to education". New York: Columbia University Press.

[6] Caroleo, F. E., Pastore, F. (2016). "Overeducation: A disease of the school-to-work transition system", in Coppola, G.L., O'Higgins, N. (eds.), *Youth and their future. Unemployment, education and health in Europe*, Routledge, London and New York.

[7] Centeno, M., Novo, A. (2012). "Excess Worker Turnover and Fixed-Term Contracts: Causal Evidence in a Two-Tier System". Labour Economics, 19(3): 320-328.

[8] Chevalier, A. (2003). "Measuring over-education". Economica, 70(279): 509-531.

[9] Chevalier, A., Lindley, J. (2009). "Overeducation and the skills of UK graduates". Journal of the Royal Statistical Society, 172(2): 307-337.

[10] Clark, B., Joubert, C., Maurel, A. (2017). "The career prospects of overeducated Americans". IZA Journal of Labor Economics, 6(3): 1-29.

[11] Cohn, E., Khan, S. P. (1995). "The wage effect of overschooling revisited". Labour Economics, 2(1): 67-76.

[12] Cohn, E., Ng, Y. C. (2000). "Incidence and wage effects of overschooling and underschooling in Hong Kong". Economics of Education Review, 19 (2): 159-168.

[13] Correia, S. (2016). "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects". Statistical Software Components s457874, Boston College Department of Economics, revised 25 Jul 2015.

[14] Cutillo, A., Di Pietro, G. (2006). "The effects of overeducation on wages in Italy: A Bivariate selectivity approach". International Journal of Manpower, 27(2): 143-168.

[15] Dahl, M. S., Klepper, S. (2015). "Whom do new firms hire?". Industrial and Corporate Change, 24(4): 819-836.

[16] Dolton, P., Vignoles. A. (2000). "The incidence and effects of overeducation in the U.K. graduate labor market". Economics of Education Review, 19(2): 179-198.

[17] Duncan, G. J, Hoffman, S. D. (1981). "The incidence and wage effects of overeducation". Economics of Education Review, 1(1): 75-86.

[18] Erdogan, B., Bauer, T. N., Peiró, J. M., Truxillo, D. M. (2011). "Overqualified employees: Making the best of a potential bad situation for individuals and organizations". Industrial and Organizational Psychology, 4(2): 215-232.

[19] Figueiredo, H., Teixeira, P., Rubery, J. (2013). "Unequal futures? Mass higher education and graduates' relative earnings in Portugal, 1995-2009". Applied Economics Letters, 20(10): 991-997.

[20] Fine, S. (2007). "Overqualification and selection in leadership training". Journal of Leadership and Organizational Studies, 14(1): 61-68.

[21] Fine, S., Nevo, B. (2008). "Too smart for their own good? A study of perceived cognitive overqualification in workforce". The International Journal of Human Resource Management, 19(2): 346-355.

[22] Fonseca, T., Lima, F., Pereira, S. (2018). "Job polarization, technological change and routinization: Evidence for Portugal". Labour Economics, 51: 317-339.

[23] Frenette, M. (2004). "The overqualified Canadian graduate: The role of the academic program in the incidence, persistence, and economic returns to overqualification". Economics of Education Review, 23(1): 29-45.

[24] Gelbach, J. (2016). "When do covariates matter? And which ones, and how much?" Journal of Labor Economics, 34(2): 509–543.

[25] Greene, F., McIntosh, S. (2007). "Is there a genuine under-utilization of skills amongst the over-qualified?". Applied Economics, 39(4): 427-439.

[26] Groot, W. (1993). "Overeducation and the returns to enterprise-related schooling". Economics of Education Review, 12(4): 299-309.

[27] Groot, W., van den Brink M. H. (2000). "Overeducation in the labor market: A meta-analysis". Economics of Education Review, 19(2): 149-158.

[28] Guimarães. P., Portugal, P. (2010). "A Simple Feasible Alternative Procedure to Fit Models with High-Dimensional Fixed Effects". The Stata Journal, 10(4): 628-49.

[29] Hartog, J. (2000). "Over-education and earnings: Where are we, where should we go?". Economics of Education Review, 19(2): 131-147.

[30] Hartog, J., Oosterbeek, H. (1988). "Education, allocation and earnings in the Netherlands: Overschooling?". Economics of Education Review, 7(2): 185-194.

[31] Hartog, J., Groeneveld, S. (2004). "Overeducation, wages and promotions within the firm". Labour Economics, 11(6): 701-714.

[32] Hersch, J. (1991). "Education match and job match". Review of Economics and Statistics, 73(1): 140-144.

[33] ILO (2014). "Skills mismatch in Europe: statistics brief". International Labour Office (Geneva).

[34] Iranzo, S., Shivardi, F., Tosetti, E. (2008). "Skill dispersion and firm productivity: An analysis with employer-employee data". Journal of Labor Economics, 26(2): 247-285.

[35] Jacobson, L. S., LaLonde, R. J., Sullivan, D. G. (1993). "Earnings Losses of Displaced Workers". The American Economic Review, 83(4): 685-709.

[36] Kampelmann, S., Mahy, B., Rycx, F., Vermeylen, G. (2020). "Over-, Required, and Undereducation: Consequences on the Bottom Lines of Firms". Labour, 34(1): 80-112

[37] Kiker, B. F., Santos, M. C., De Oliveira, M. M. (1997). "Overeducation and undereducation: evidence for Portugal". Economics of Education Review, 16(2): 111-125.

[38] Korpi, T.; Tåhlin, M. (2009). "Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000". Labour Economics, 16(2): 183-193.

[39] Martins, P. (2009). "Dismissals for cause: The difference that just eight paragraphs can make". Journal of Labor Economics, 27(2): 257-279.

[40] Maynard, D. C., Joseph, T. A., Maynard, A. M. (2006). "Underemployment, job attitudes, and turnover intensions". Journal of Organizational Behavior, 27(4): 509-536.

[41] Mavromaras, K., McGuinness, S., O'Leary, N., Sloane, P, Wei, Z. (2013). "Job mismatches and labour market outcomes: Panel evidence on Australian university graduates". Economic Record, 89 (286): 382-395.

[42] McGuinness, S. (2006). "Overeducation in the labor market?". Journal of Economic Surveys, 20(3): 387-418.

[43] McGuinness, S., Pouliakas, K., Redmond, P. (2018). "Skills mismatch: Concepts, measurement and policy approaches". Journal of Economic Surveys, 32(4): 985-1015.

[44] Murillo, I. P., Rahona-López, M., Salinas-Jiménez, M. M. (2012). "Effects of educational mismatch on private returns to education: An analysis of the Spanish case (1995-2006)". Journal of Policy Modeling, 34(5): 646-659.

[45] Mysíkovám, M. (2016). "Educational mismatch in the Czech labour market". Review of Economic Perspectives, 16(2): 103-120.

[46] OECD (2012). OECD Economic Surveys: Portugal 2012. OECD Publishing, Paris.

[47] OECD (2017). Education at a glance 2017: OECD Indicators. OECD Publishing, Paris.

[48] Oliveira, M. M., Santos, M. C., Kiker, B. F. (2000), "The role of human capital and technological change in overeducation". Economics of Education Review, 19(2): 199-206.

[49] Quintini, G. (2011). "Over-qualified or under-skilled: A review of existing literature". OECD Social, Employment and Migration WP No 121. DELSA/ELSA/WD/SEM(2011)6.

[50] Rocha, V., van Praag, M., Folta, T., Carneiro, A. (2019). "Endogeneity in strategy-performance analysis: An application to initial human capital strategy and new venture performance". Organizational Research Methods, 22(3): 740-764.

[51] Rumberger, R.W. (1987). "The impact of surplus schooling on productivity and earnings". Journal of Human Resources, 22(1): 24-50.

[52] Sala, G. (2011). “Approaches to Skills Mismatch in the Labour Market: A Literature Review”. Papers: Revista de Sociologia, 96(4), 1025–1045.

[53] Sattinger, M. (1993). "Assignment models of the distribution of earnings". Journal of Economic Literature, 31(2): 831-880.

[54] Sicherman, N. (1991). "Overeducation in the labor market". Journal of Labor Economics, 9(2): 101-122.

[55] Tavares, M, Carneiro, A, Varejão, J. (2018). "The spatial dimension of internal labor markets". Journal of Regional Science, 58: 181-203.

[56] Thurow, L. C. (1975). "A job-competition model", In: Priore, M. J. (Ed.): *Unemployment and inflation: Institutionalist and Structuralist Views*. New York, 17-32.

[57] Tsai, Y. (2010). "Returns to overeducation: A longitudinal analysis of the U.S. labor market". Economics of Education Review, 29(4): 606-617.

[58] Tsang, M. C., Levin, H. M. (1985). "The Economics of Overeducation ". Economics of Education Review, 4(2): 93-104.

[59] Tsang, M. C., Rumberger, R. W., Levin, H. M. (1991). "The impact of surplus schooling on worker productivity". Industrial Relations, 30(2): 209-228.

[60] Van Smoorenburg, M., Van der Velden, R. (2000). "The training of school leavers, complementarity or substitution?". Economics of Education Review, 19(2): 207-217.

[61] Verdugo, R., Verdugo, N. (1989). "The impact of surplus schooling on earnings". Journal of Human Resources, 24(4): 629-643.

TABLES AND FIGURES

TABLE 1: Number of observations by group of educational mismatch at job entry, Portugal 1998-2012

Year	Overeducated (OE)	Undereducated (UE)	Matched (M)	Total
t (baseline year)	1,018,476	795,101	1,579,804	3,393,381
$t+1$	523,677	421,071	814,506	1,759,254
$t+2$	370,403	285,656	570,507	1,226,566
$t+3$	250,240	191,475	383,639	825,354
Total	2,253,564	1,801,933	3,489,618	7,204,555

Note: For each of the three columns, the number of observations in the first row indicates the number of workers that in the year of being hired were OE, UE, or M. The numbers in the other rows indicate the number of those individuals that 1, 2, or 3 years ahead survive in the original firm. These numbers decrease over time as those workers move to a new firm and the individual clock resets to zero or due to exit from the QP files.

TABLE 2: Incidence of overeducation and undereducation for newly hired workers, Portugal 1998-2012

	Full sample	Males	Females
	N=3,393,381	N=2,077,467	N=1,315,914
Overeducated (%)	30.0	31.6	27.5
Undereducated (%)	23.4	23.9	22.6
Matched (%)	46.6	44.5	49.9
Total	100.0	100.0	100.0

TABLE 3: Incidence of over- and undereducation by ccupation for newly hired workers, Portugal 1998-2012

Major group	Occupation (1-digit)	Number of observations	OE (%)	UE (%)	M (%)
1	Managers	45,869	21.9	36.0	42.1
2	Professionals	273,472	2.6	26.8	70.6
3	Technicians and associate professionals	301,950	26.8	29.9	43.4
4	Clerical support workers	483,921	23.4	24.1	52.5
5	Service and sales workers	294,814	26.1	30.1	43.8
6	Skilled agricultural, forestry and fishery workers	66,455	40.8	9.0	50.2
7	Craft and related trade workers	804,818	36.8	21.5	41.7
8	Plant and machine operators and assemblers	460,008	29.4	29.0	41.6
9	Elementary occupations	662,074	41.1	14.7	44.2
	Total	3,393,381			

TABLE 4: Real hourly earnings (in logs) by group of educational mismatch at job entry, Portugal 1998-2012

Year	Overeducated (OE)	Undereducated (UE)	Matched (M)
t (baseline year)	1.375	1.365	1.426
$t+1$	1.431	1.416	1.484
$t+2$	1.477	1.461	1.525
$t+3$	1.523	1.498	1.560

TABLE 5: Descriptive statistics, Portugal 1998-2012
$N = 7,204,555$

	Mean	St. Dev.	Min.	Max.
Log real hourly earnings (in 2010 euros)	1.44	0.45	-0.38	6.52
Female	0.39		0	1
Age (years)	33.92	10.15	17	65
Education (years)	8.79	4.06	0	16
Tenure (months)	16.25	13.24	0	47
Qualification levels				
Top executives	0.05		0	1
Intermediary executives	0.04		0	1
Supervisor, team leader, foreman	0.03		0	1
High-skilled professionals	0.06		0	1
Skilled professionals	0.39		0	1
Semi-skilled professionals	0.17		0	1
Non-skilled professionals	0.17		0	1
Apprentices, interns, trainees	0.06		0	1
Non-defined	0.04		0	1

TABLE 6: Wage regressions results, Portugal 1998-2012
Dependent variable: log real hourly wages

	OLS Base Model	Worker FE	Worker & Firm FE Full Model
	(1)	(2)	(3)
OE_0	−0.0734***	−0.0173***	−0.0159***
	(0.00048)	(0.00063)	(0.00057)
OE_1	−0.0687***	−0.0171***	−0.0156***
	(0.00057)	(0.00069)	(0.00061)
OE_2	−0.0580***	−0.0162***	−0.0142***
	(0.00067)	(0.00074)	(0.00066)
OE_3	−0.0473***	−0.0092***	−0.0077***
	(0.00087)	(0.00082)	(0.00074)
UE_0	0.0580***	0.0082***	0.0082***
	(0.00051)	(0.00066)	(0.00059)
UE_1	0.0503***	0.0035***	0.0056***
	(0.00063)	(0.00073)	(0.00065)
UE_2	0.0473***	0.0010	0.0043***
	(0.00076)	(0.00079)	(0.00070)
UE_3	0.0449***	0.0046***	0.0078***
	(0.00096)	(0.00088)	(0.00078)
R^2 *overall*	0.47	0.84	0.89
N	7,204,555	7,204,555	7,204,555

Notes: (i) The control variables include age, age^2, tenure, tenure^2, gender dummy, qualification dummies, and time dummies;
(ii) worker-cluster robust standard errors in parentheses;
(iii) ***, **, * denote significant at 1, 5, and 10 percent, respectively.

TABLE 7: Decomposition of the change in the wage gap of the over- and underducated workers

	Base Model Coeffic. estimate (1)	Full Model Coeffic. estimate (2)	Change (1)-(2) (3)	Decomposition of the change into: Worker FE (4)	Firm FE (5)
OE_0	−0.0734	−0.0159	−0.0575	−0.0334	−0.0241
				(0.00029)	(0.00026)
OE_1	−0.0687	−0.0156	−0.0531	−0.0331	−0.0201
				(0.000 36)	(0.00032)
OE_2	−0.0580	−0.0142	−0.0438	−0.0290	−0.0148
				(0.00042)	(0.00037)
OE_3	−0.0473	−0.0077	−0.0396	−0.0274	−0.0122
				(0.00055)	(0.00049)
UE_0	0.0580	0.0082	0.0498	0 .0373	0.0126
				(0.00031)	(0.00028)
UE_1	0.0503	0.0056	0.0447	0.0352	0.0095
				(0.00039)	(0.00035)
UE_2	0.0473	0.0043	0.0430	0.0337	0.0094
				(0.00047)	(0.00042)
UE_3	0.0449	0.0078	0.0371	0.0302	0.0069
				(0.00061)	(0.00054)

Notes: This table reports the decomposition of the wage gap variation for mismatched workers from the base (column 1) to the full models (column 3) of Table 6. The figures in columns (1) and (2) of this Table are the estimates of the coefficients of the OE_k and UE_k dummies for the base and full models, respectively. Column (3) reports the difference in the estimates of the mismatched dummies between columns (1) and (2). Columns (4) and (5) present the contribution of the corresponding fixed effect for the observed change in the estimates of the mismatched dummies from the base to the the full model, computed according to the Gelbach procedure. Robust worker cluster standard errors in parentheses.

TABLE 8: Wage regressions results (alternative wage measure), Portugal 1998-2012
Dependent variable: log real hourly base wages

	OLS	Worker FE	Worker & Firm FE
	(1)	(2)	(3)
OE_0	−0.0732***	−0.0161***	−0.0142***
	(0.00043)	(0.00056)	(0.00052)
OE_1	−0.0637***	− 0.0142***	−0.0130***
	(0.00051)	(0.00061)	(0.00055)
OE_2	−0.0548***	−0.0135***	−0.0118***
	(0.00061)	(0.00065)	(0.00059)
OE_3	−0.0477***	− 0.0097***	−0.0079***
	(0.00078)	(0.00072)	(0.00066)
UE_0	0.0559***	0.0072***	0.0060***
	(0.00046)	(0.00057)	(0.00053)
UE_1	0.0518***	0.0032***	0.0037***
	(0.00056)	(0.00064)	(0.00057)
UE_2	0.0465***	0.0012*	0.0025***
	(0.00068)	(0.00069)	(0.00062)
UE_3	0.0411***	0.0022***	0.0037***
	(0.00087)	(0.00077)	(0.00069)
R^2 *overall*	0.53	0.87	0.91
N	7,204,555	7,204,555	7,204,555

Notes: (i) The control variables include age, age^2, tenure, tenure^2, gender dummy, qualification dummies, and time dummies;
(ii) worker-cluster robust standard errors in parentheses;
(iii) ***, **, * denote significant at 1, 5, and 10 percent, respectively.

TABLE 9: Wage regressions results (including interaction terms with experience), Portugal 1998-2012

Dependent variable: log real hourly wages

	OLS	Worker FE	Worker & Firm FE
	(1)	(2)	(3)
OE_0	−0.0895***	−0.0570***	− 0.0514***
	(0.00082)	(0.0010)	(0.00089)
OE_1	−0.0711***	−0.0485***	− 0.0425***
	(0.0011)	(0.0012)	(0.0010)
OE_2	−0.0574***	−0.0386***	−0.0300***
	(0.0016)	(0.0014)	(0.0012)
OE_3	−0.0444***	−0.0232***	−0.0117***
	(0.0025)	(0.0019)	(0.0018)
UE_0	−0.0485***	−0.0478***	−0.0336***
	(0.0020)	(0.0021)	(0.0019)
UE_1	−0.0605***	−0.0510***	−0.0281***
	(0.0030)	(0.0026)	(0.0023)
UE_2	−0.0700***	−0.0415***	−0.0130***
	(0.0045)	(0.0035)	(0.0030)
UE_3	−0.0601***	−0.0211***	0.0108**
	(0.0065)	(0.0046)	(0.0042)
$OE_0 * \exp$	0.0167***	0.0463***	0.0417***
	(0.00084)	(0.00092)	(0.00080)
$OE_1 * \exp$	0.0011	0.0337***	0.0291***
	(0.0011)	(0.0011)	(0.00091)
$OE_2 * \exp$	−0.0034**	0.0217***	0.0150***
	(0.0016)	(0.0013)	(0.0011)
$OE_3 * \exp$	−0.0048*	0.0117***	0.0016
	(0.0026)	(0.0018)	(0.0017)
$UE_0 * \exp$	0.1132***	0.0598***	0.0448***
	(0.0021)	(0.0021)	(0.0018)
$UE_1 * \exp$	0.1183***	0.0575***	0.0360***
	(0.0030)	(0.0026)	(0.0022)
$UE_2 * \exp$	0.1211***	0.0445***	0.0185***
	(0.0045)	(0.0035)	(0.0030)
$UE_3 * \exp$	0.1086***	0.0267***	−0.0024
	(0.0065)	(0.0046)	(0.0042)
R^2 *overall*	0.47	0.84	0.89
N	7,204,555	7,204,555	7,204,555

Notes: see notes to Table 6. $exp = 1\ if\ age - education - 6 > 5$.

FIGURE 1: Incidence of over- and undereducation for newly hired workers, Portugal 1998-2012

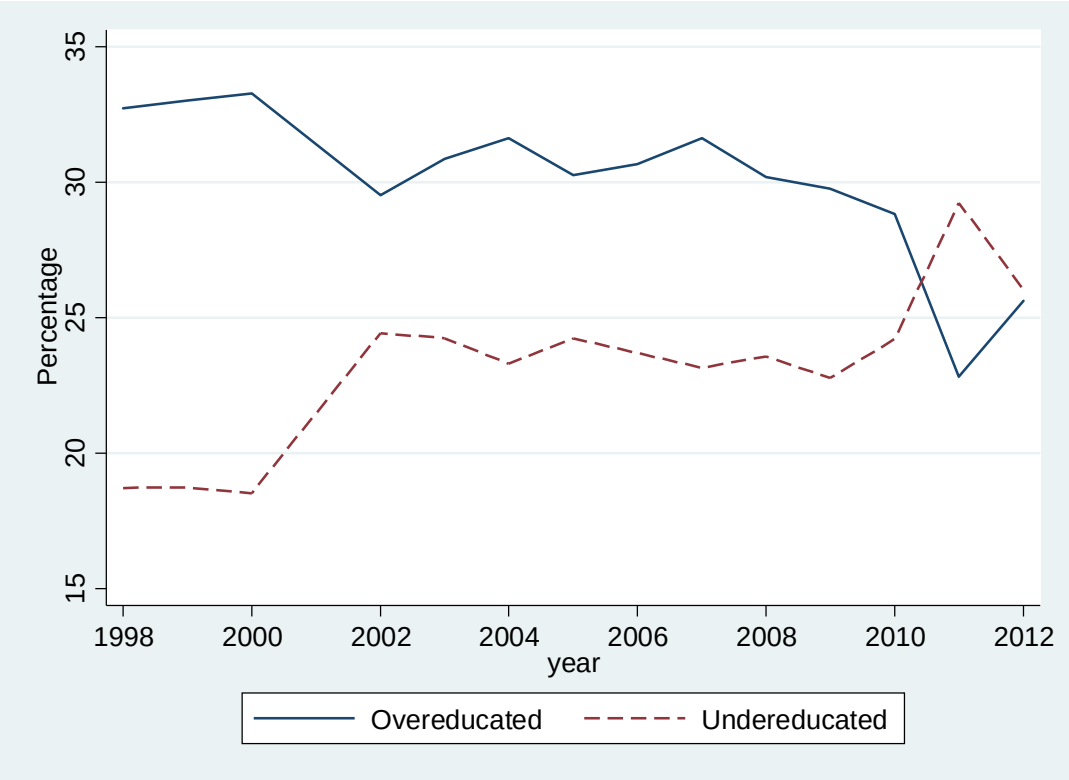

FIGURE 2: Average schooling years for newly hired workers, Portugal 1998-2012

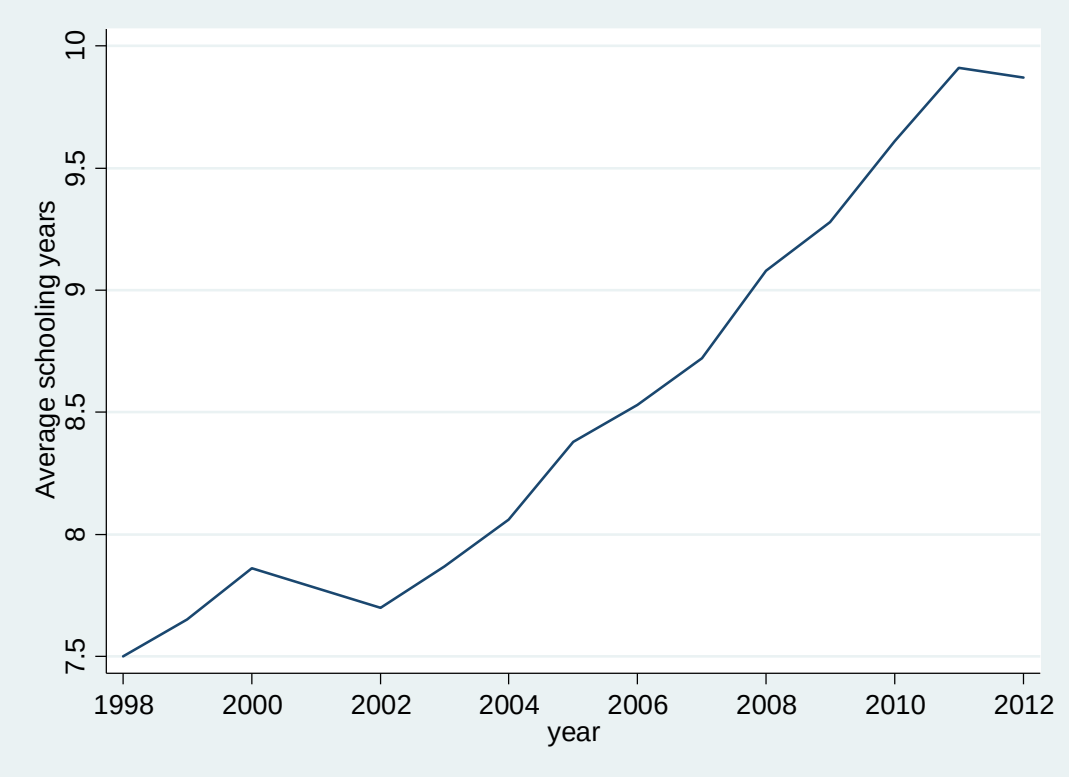

FIGURE 3: The empirical distribution of worker permanent heterogeneity for overeducated and matched workers

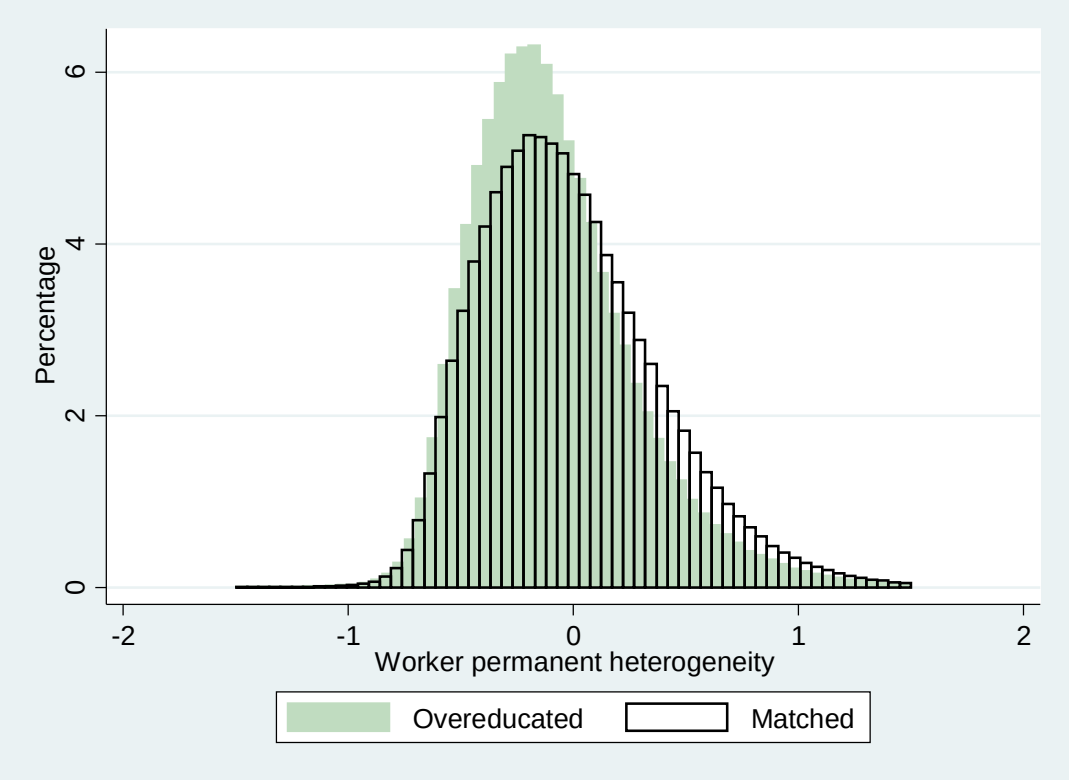

FIGURE 4: The empirical distribution of worker permanent heterogeneity for undereducated and matched workers

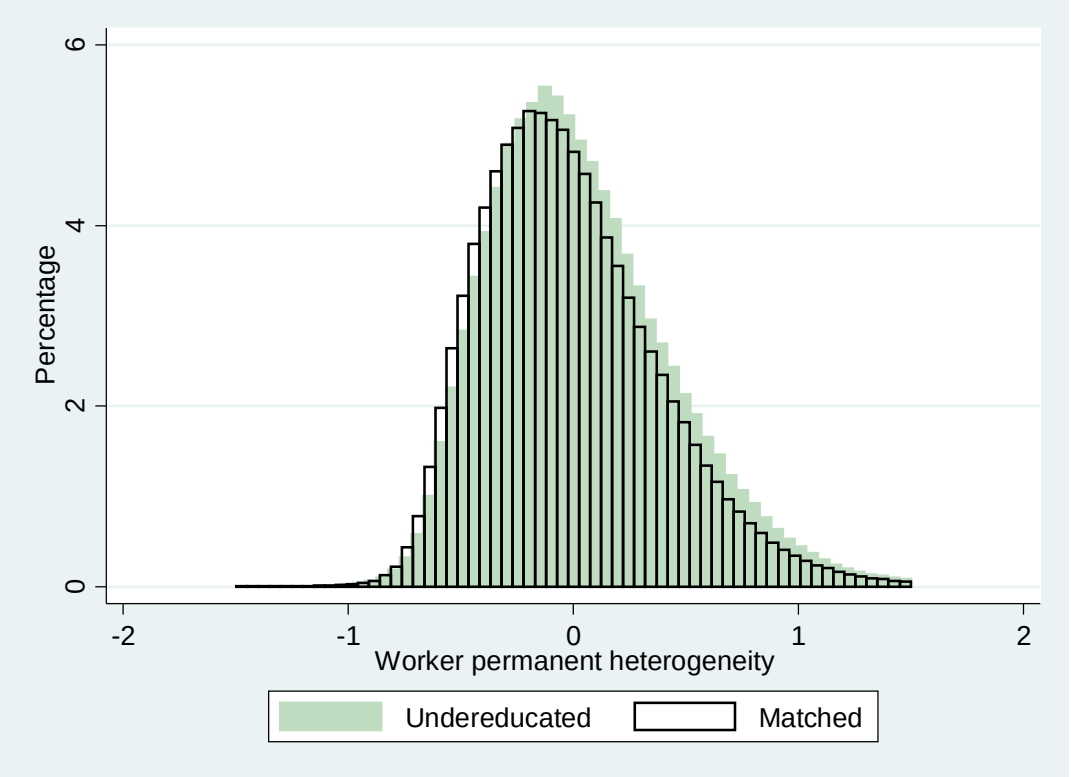

FIGURE 5: The empirical distribution of firm permanent heterogeneity for overeducated and matched workers

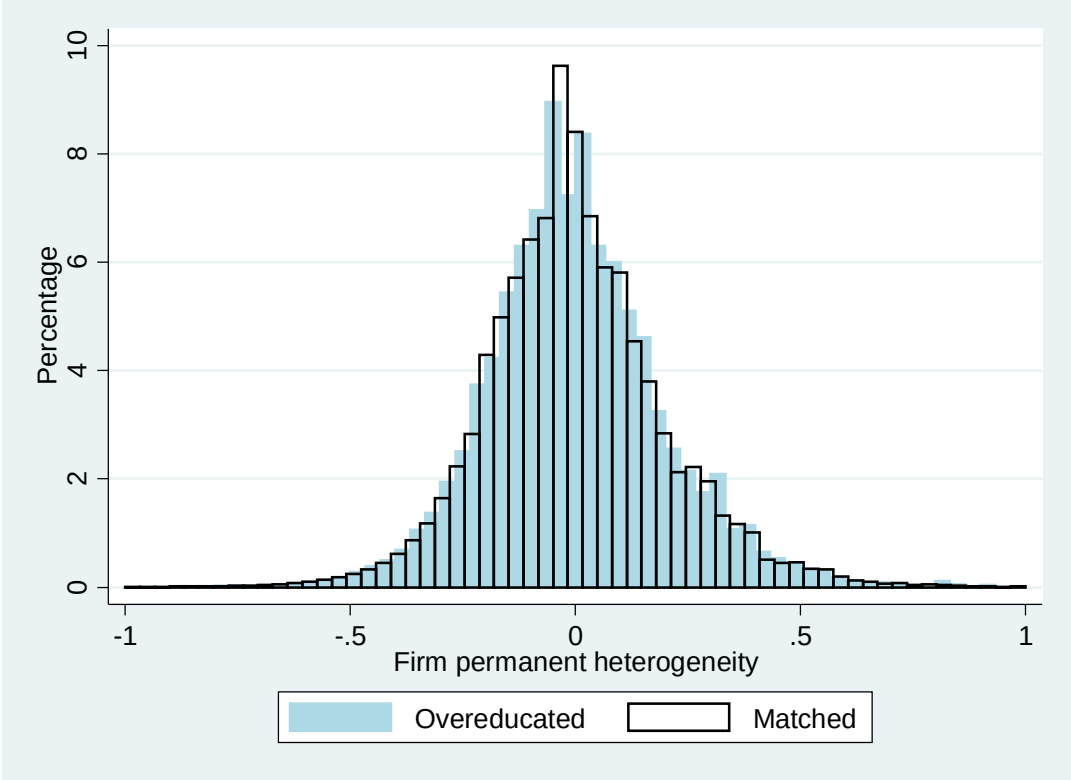

FIGURE 6: The empirical distribution of worker permanent heterogeneity for undereducated and matched workers